



Gupta, K., Sari-Ak, D., Haffke, M., Trowitzsch, S., & Berger, I. (2016). Zooming in on Transcription Preinitiation. *Journal of Molecular Biology*, 428(12), 2581-2591.
<https://doi.org/10.1016/j.jmb.2016.04.003>

Publisher's PDF, also known as Version of record

License (if available):
CC BY-NC-ND

Link to published version (if available):
[10.1016/j.jmb.2016.04.003](https://doi.org/10.1016/j.jmb.2016.04.003)

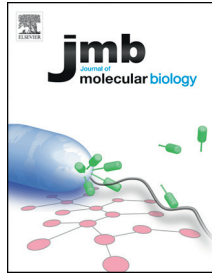
[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the corrected proof version of the article (version of record). It first appeared online via Elsevier at <http://www.sciencedirect.com/science/article/pii/S0022283616300468>.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



Zooming in on Transcription Preinitiation

Kapil Gupta^{1,2}, Duygu Sari-Ak^{1,2}, Matthias Haffke³,
Simon Trowitzsch⁴ and Imre Berger^{1,2,5}

1 - European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, 38042, Grenoble Cedex 9, France

2 - Unit of Virus Host-Cell Interactions, University of Grenoble Alpes-EMBL-CNRS, UMI 3265, 71 Avenue des Martyrs, 38042, Grenoble, Cedex 9, France

3 - Center for Proteomic Chemistry, Structural Biophysics, Novartis Institute for Biomedical Research NIBR, Fabrikstrasse 2, 4056 Basel, Switzerland

4 - Institute of Biochemistry, Biocenter, Goethe-Universität Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt/Main Germany

5 - The School of Biochemistry, University of Bristol, Bristol BS8 1TD, UK

Correspondence to Imre Berger: The School of Biochemistry, University of Bristol, University Walk, Clifton, BS8 1TD, Bristol, UK. imre.berger@bristol.ac.uk

<http://dx.doi.org/10.1016/j.jmb.2016.04.003>

Edited by Lori A Passmore

Abstract

Class II gene transcription commences with the assembly of the Preinitiation Complex (PIC) from a plethora of proteins and protein assemblies in the nucleus, including the General Transcription Factors (GTFs), RNA polymerase II (RNA pol II), co-activators, co-repressors, and more. TFIID, a megadalton-sized multiprotein complex comprising 20 subunits, is among the first GTFs to bind the core promoter. TFIID assists in nucleating PIC formation, completed by binding of further factors in a highly regulated stepwise fashion. Recent results indicate that TFIID itself is built from distinct preformed submodules, which reside in the nucleus but also in the cytosol of cells. Here, we highlight recent insights in transcription factor assembly and the regulation of transcription preinitiation.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

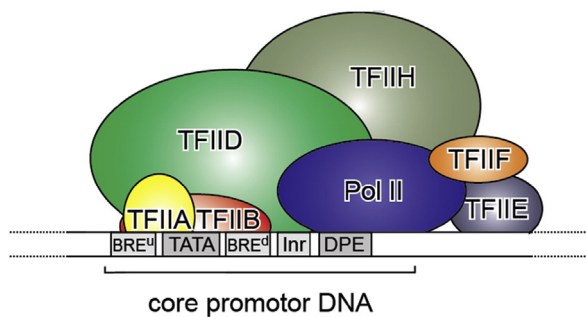
Introduction

Class II gene transcription is a tightly regulated, essential process controlled by a highly complex multicomponent machinery. A plethora of proteins, more than a hundred in humans, are organized in often very large multiprotein assemblies including General Transcription Factors (GTFs TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH), RNA polymerase (RNA pol II), and a large number of diverse complexes that act as co-activators, co-repressors, chromatin modifiers and remodelers (Fig. 1). Class II gene transcription is regulated at various levels: while assembling on chromatin, before and during transcription initiation, throughout elongation and mRNA processing, and termination. A host of activators and repressors has been reported to regulate transcription [1]. A central multisubunit complex called the Mediator

was identified as a global regulator of gene expression [2–4]. Functional and structural analyses of many components of this striking complexity have provided immense insights into the transcription process. In this contribution, we are reviewing, by no means exhaustively, recent important findings about key architectures within the transcription machinery, leading to conceptual advances in terms of complex assembly and function, with a focus on the key GTF that nucleates PIC formation, TFIID.

PIC Assembly: Lessons from Yeast and Human

Transcription of RNA pol II-dependent genes is triggered by the regulated assembly of the Preinitiation Complex (PIC). PIC formation commences with



TFIID—TFIIA—TFIIB—Pol II/TFIIF—TFIIE—TFIIH
Sequential Assembly Pathway

Fig. 1. Transcription PIC. Class II gene transcription is brought about by (in humans) over a hundred polypeptides assembling on the core promoter of protein-encoding genes, which then give rise to messenger RNA. A PIC on a core promoter is shown in a schematic representation (adapted from Ref. [5]). PIC contains, in addition to promoter DNA, the GTFs TFIIA, B, D, E, F, and H, and RNA Pol II. PIC assembly is thought to occur in a highly regulated, stepwise fashion (top). TFIID is among the first GTFs to bind the core promoter via its TBP subunit. Nucleosomes at transcription start sites contribute to PIC assembly, mediated by signaling through epigenetic marks on histone tails. The Mediator (not shown) is a further central multiprotein complex identified as a global transcriptional regulator. TATA, TATA-box DNA; BRE^u, B recognition element upstream; BRE^d, B recognition element downstream; Inr, Initiator; DPE, Down-stream promoter element.

the binding of TFIID to the core promoter. TFIID contains the TATA-box binding protein (TBP). Binding of TFIID to the core-promoter is followed by the recruitment of further GTFs and RNA pol II. Several lines of evidence suggest that this process occurs in a defined, stepwise order and undergoes significant restructuring [5]. First, PIC adopts an inactive state, the “closed” complex, which is incompetent to initiate transcription. The ATP-dependent translocase activity of the XPB/Xpd helicase subunit of GTF TFIIH then opens up about 11 to 15 base pairs around the transcription start site by moving along one DNA strand inducing torsional strain, leading to conformational rearrangements and positioning of single-stranded DNA to the active site of RNA pol II [6–9]. In this “open” complex, RNA pol II can enter elongation to transcribe throughout a gene in a highly processive manner without dissociating from the DNA template or losing the nascent RNA. In most eukaryotes, after synthesizing about 20–100 bases, RNA pol II can pause (Promoter proximal pause) and then disconnect from promoter elements and other components of the transcription machinery, giving rise to a fully functional elongation complex in a process called promoter escape [10–14]. The promoter-bound components of the PIC, in contrast, remain in place, and thus only TFIIB, TFIIF, and RNA pol II need to be recruited for re-initiation, significantly increasing the

transcription rate in subsequent rounds of transcription [12,15]. Promoter escape is preceded by an abortive transcription in many systems, where multiple short RNA products of 3 to 10 bases in length are synthesized [16,17].

Recent landmark studies on human and yeast PIC formation provided more differentiated views of the first steps in the transcription initiation process, corroborating the concept of stepwise assembly while also hinting at significant differences that may be present between the species [18,19] (reviewed in Ref. [20]). In the study of the human PIC, the proposed assembly mechanism follows the “conventional” stepwise order with the exception that RNA pol II appears to be already recruited at the very beginning, before TFIIF is accreted [18]. According to this model, TFIIF functions in reorganizing the growing PIC, rather than loading RNA pol II into it. TFIIH is the last component to be recruited [18]. According to the model put forward based on the studies from yeast, all GTFs (except TFIIF) including TFIIH assemble into a PIC lacking RNA pol II, which, together with TFIIF, is the last to be incorporated [19]. The structures and models presented in these ground-breaking reports provide a wealth of architectural and functional insight into PIC assembly and convey that there may be different ways to organize PIC in space and time, and that differences between the species may exist [18,19]. In a separate study, the architecture of a yeast initially transcribing complex (ITC) was determined [21]. ITC is an intermediate complex formed during PIC assembly by RNA pol II, TFIIF, TFIIB, TBP, and DNA, as well as a small nascent RNA [21]. Interestingly, this study revealed similarities with the model of human ITC [18], suggesting that the core architecture of PIC is conserved between yeast and human.

Notably, the described studies above used TBP instead of holo-TFIID. TBP has been shown to suffice for basal transcription, whereas holo-TFIID is required for activated transcription [22,23]. Therefore, it is conceivable that PIC assembly may follow alternative pathways in activated transcription.

Mediator Core–RNA Pol II Initially Transcribing Complex

Recently, single particle cryo-electron microscopy (cryo-EM) and crosslinking mass spectrometry (CLMS) studies of a yeast ITC bound to a Mediator core complex revealed important first insights into transcription initiation and PIC assembly, suggesting that Mediator is involved in stabilizing PIC and in activating RNA pol II [24]. The architecture of Mediator-ITC is shown in Fig. 2. The cryo-EM structure was determined at nanometer resolution, and CLMS proved to be instrumental to decipher the subunit topology of the Mediator Middle module.

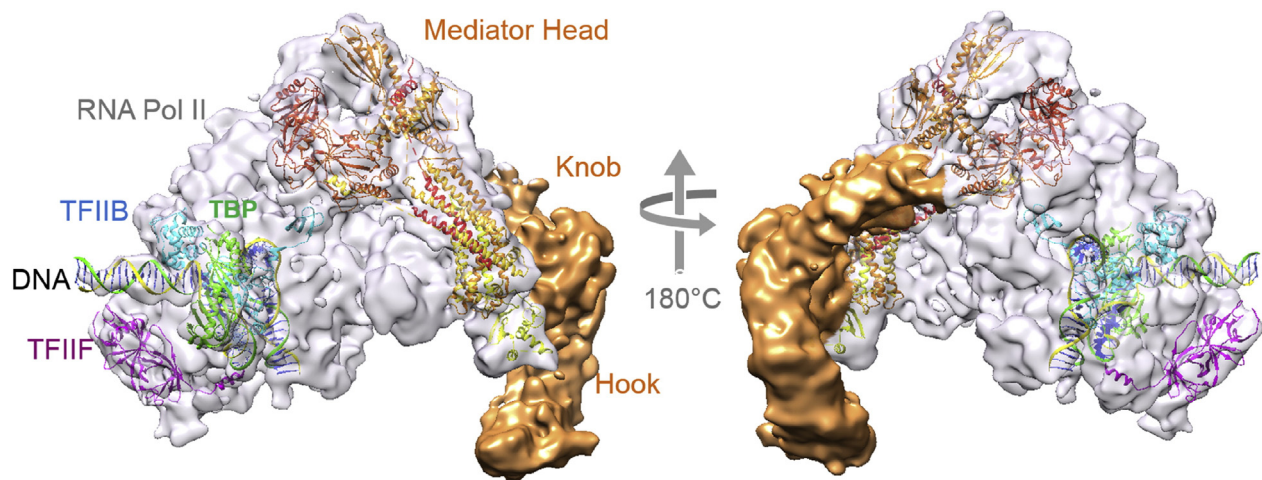


Fig. 2. Mediator Core–ITC Architecture. The structure of a Mediator core bound to an initially transcribing RNA pol II-containing complex (EMD-2786; PDB ID: 4V1O) is shown in two views related by a 180° rotation along the vertical axis (arrow). TBP (green), promoter DNA (yellow and green), TFIIB (blue), TFIIF (purple), and the mediator Head module are depicted in a cartoon representation based on X-ray crystal structure coordinates. RNA Pol II is colored in gray. Mediator Middle Module (Knob, Hook) is colored in orange. The structure was determined by cryo-EM combined with CLMS and by fitting of available atomic coordinates (adapted from Ref. [24]).

Mediator interacts with TFIIB and RNA pol II, engaging three distinct interfaces, one of which is thought to be more transient. One interface features an interaction of the so-called mobile jaw of the Mediator Head with the β -ribbon domain of TFIIB and RNA pol II. The second interface is formed by the conserved arm domain of the Head engaging RNA pol II. *In vivo* experiments showed reduced mRNA synthesis when the observed Mediator–TFIIB–RNA pol II interactions were perturbed, confirming the importance of the identified interfaces [20]. These results provide a molecular rationale for the key roles of Mediator in recruiting TFIIB, stabilizing PIC, and activating RNA pol II.

TFIID Structural Plasticity

A key component in transcription initiation is GTF TFIID, a large megadalton-sized multiprotein complex with around 20 subunits made up of 14 different polypeptides: TBP and the TBP-associated factors (TAFs) (numbered 1–13) (Fig. 3) [25]. Analyses of subunit stoichiometry within TFIID revealed that a number of TAFs are present in two copies while others are found in single copy (Fig. 3) [26]. A key feature in TAFs is the histone fold domain (HFD), which is present in 9 out of 13 TAFs in TFIID. The HFD is a strong protein–protein interaction motif that mediates specific dimerization. The HFD-containing TAFs are organized in discrete heterodimers, with the exception of TAF10, which is capable of forming dimers with two different TFIID components, TAF3 and TAF8. HFDs and several other structural features of TBP and the TAFs are well conserved between the species [27,28].

TFIID was shown to adopt an asymmetric, horse-shoe shape with three almost equal-sized lobes (A, B, and C), exhibiting a considerable degree of conformational flexibility with at least two distinct conformations (open and closed) [29–32]. More recent cryo-EM analyses of TFIID in the presence of TFIIA and a synthetic “super core promoter” containing a non-natural combination of promoter elements revealed a novel, completely reorganized conformational state of TFIID (Fig. 3) [33]. In this reorganized state, an entire lobe of TFIID appears to reorganize and migrate to a different position within the holo-complex. A significant number of interactions must be disrupted and reformed in such a large-scale movement, challenging the classical view of TFIID (and similar complexes) as rigidly structured assemblies.

Nuclear TFIID Core Complex

Further support for the remarkable structural flexibility of TFIID comes from recent studies of a physiological TFIID core complex. Core-TFIID consists of a subset of TAFs, TAF4, TAF5, TAF6, TAF9, and TAF12, and was first identified in *Drosophila* nuclei [34]. Hybrid studies of human core-TFIID integrating cryo-EM, data from X-ray crystallography, and homology models and proteomics revealed a twofold symmetric, intertwined architecture with large solvent channels, formed by two copies each of the constituent TAFs (Fig. 3) [35]. Interestingly, this symmetric structure was shown to reorganize into an asymmetric shape upon binding of a single TAF8–TAF10 heterodimer. The TAF8–TAF10

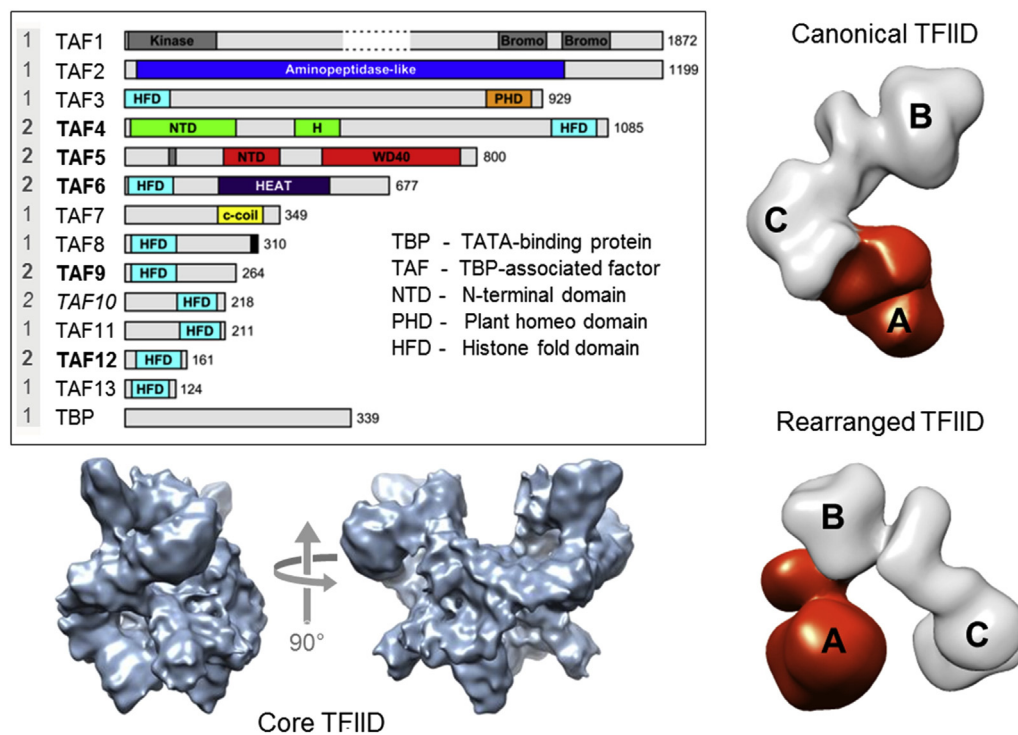


Fig. 3. Human GTF TFIID. TFIID is a large megadalton-sized multiprotein complex comprising about 20 subunits made up of 14 different polypeptides. The constituent proteins of TFIID, TBP and the TAFs, are shown in a schematic representation depicted as bars (inset, left). Structured domains are marked and annotated. The presumed stoichiometry of TAFs and TBP in the TFIID holo-complex is given (far left, gray underlined). TAF10 (in italics) makes histone fold pair separately with both TAF3 and TAF8. TAFs present in a physiological TFIID core complex extracted from eukaryotic nuclei are labeled in bold. The architecture of TFIID core complex (EMD-2230) determined by cryo-EM is shown (bottom left) in two views related by a 90° rotation (arrows) [35]. The holo-TFIID complex is characterized by remarkable structural plasticity. Two conformations, based on cryo-EM data (EMD-2284 and EMD-2287), are shown on the right, a canonical form (top) and a more recently observed rearranged form (bottom). In the rearranged conformation, lobe A (colored in red) migrates from one extreme end of the TFIID complex (attached to lobe C) all the way to the other extremity (attached to lobe B) [33].

complex binds in close vicinity to the twofold axis relating the two identical halves of core-TFIID, thereby obstructing the integration of the second copy of TAF8–TAF10 by steric hindrance. These structural rearrangements provide compelling evidence for significant reorganizations in the process of TFIID assembly, which may well channel into large-scale restructuring of the holo-complex when binding to core-promoter DNA and, possibly, other components of the preinitiation machinery.

Preformed TFIID Submodule in the Cytoplasm

Not much is known to date about the mechanisms of TFIID assembly *in vivo*. Evidence suggests that TFIID formation is a unique and stepwise process, not just a random accretion of different subunits [36]. The classical view of TFIID (and by analogy, also other multiprotein complexes) as a rigidly crafted, “canonical” complex has suffered significant erosion

recently. Tissue-specific forms of TFIID have been described, containing distinct isoforms of certain subunits [37]. Partial TFIID complexes containing only subsets of TAFs, such as, for example, nuclear core-TFIID, have been identified *in vivo*, and their unique roles in transcription regulation have been postulated [34,37–42]. Data from affinity capture experiments followed by mass-spectrometry indicate that only a fraction of the TFIID material in cells comprises a full complement of TAFs and TBP, while the majority appears to exist in partial complexes containing only a subset of TAFs [43]. Moreover, TAFs are not confined to TFIID—they are also found in the Spt-Ada-Gcn5-Acetyl transferase (SAGA) complex, which is another transcriptional co-activator of similar complexity [44]. Taken together, these findings convey a fluid situation in the nuclei of cells, in which partial TFIID and SAGA complexes may coexist with complete holo-TFIID. However, the organizational and mechanistic details of the underlying dynamics and the possible physiological consequences remain enigmatic.

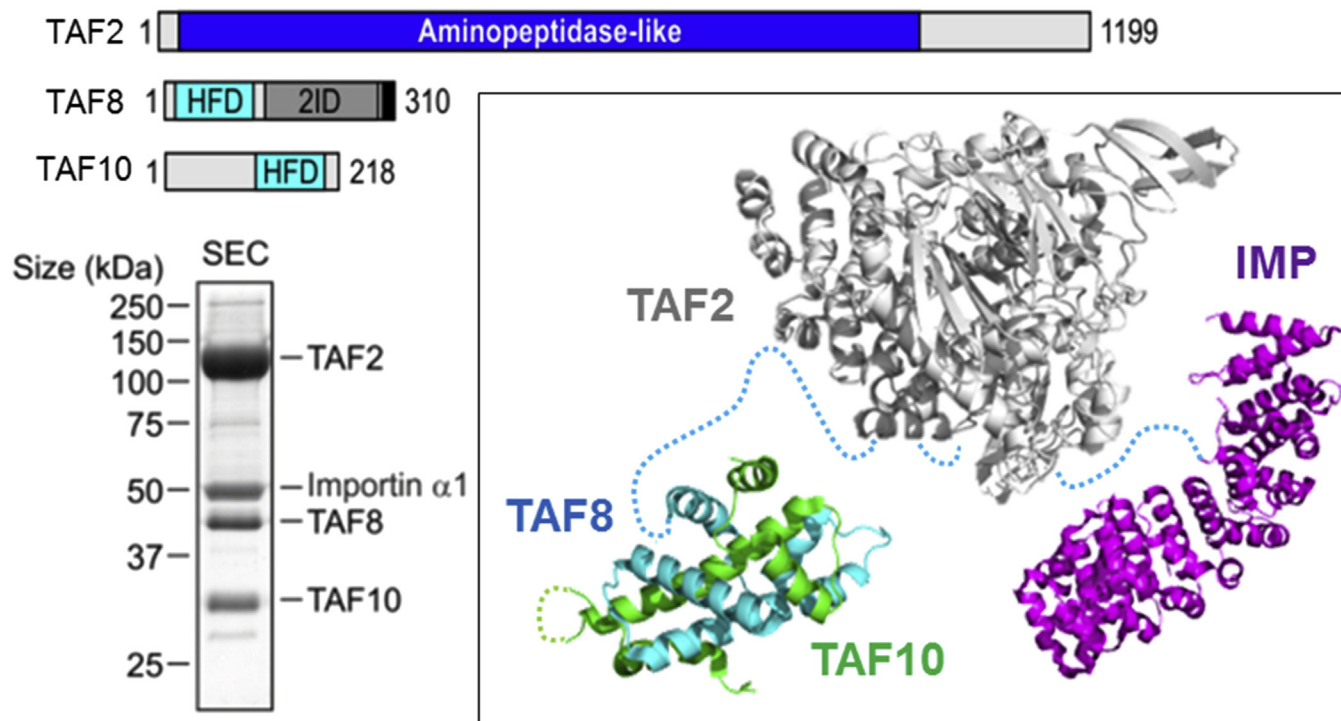


Fig. 4. Preformed TFIID Submodule in Cytoplasm. Recently, a stable TFIID subcomplex consisting of TAF2, TAF8, and TAF10 has been discovered, surprisingly residing in the cytoplasm of cells. TAF2, TAF8, and TAF10 are shown in a schematic representation (top), depicted as bars. TAF2 is characterized by an extended aminopeptidase-like fold. TAF8 and TAF10 contain HFDs. The TAF8 domain mediating interaction with TAF2 has been identified (marked as 2ID). TAF8 and TAF10 form an HFD pair in the TAF2–8–10 complex. TAF8 contains a nuclear localization signal (black box), and a stable putative nuclear import complex comprising one copy each of TAF2, TAF8, TAF10, and Importin- α could be purified to homogeneity (bottom left). A model of this NIP is shown (inset), based on crystal structures (TAF8–TAF10; TAF8–Imp) (PDB IDs: 4WV4 and 4WV6) and a model of TAF2 threaded on highly homologous human aminopeptidase ERAP (adapted from Ref. [45]).

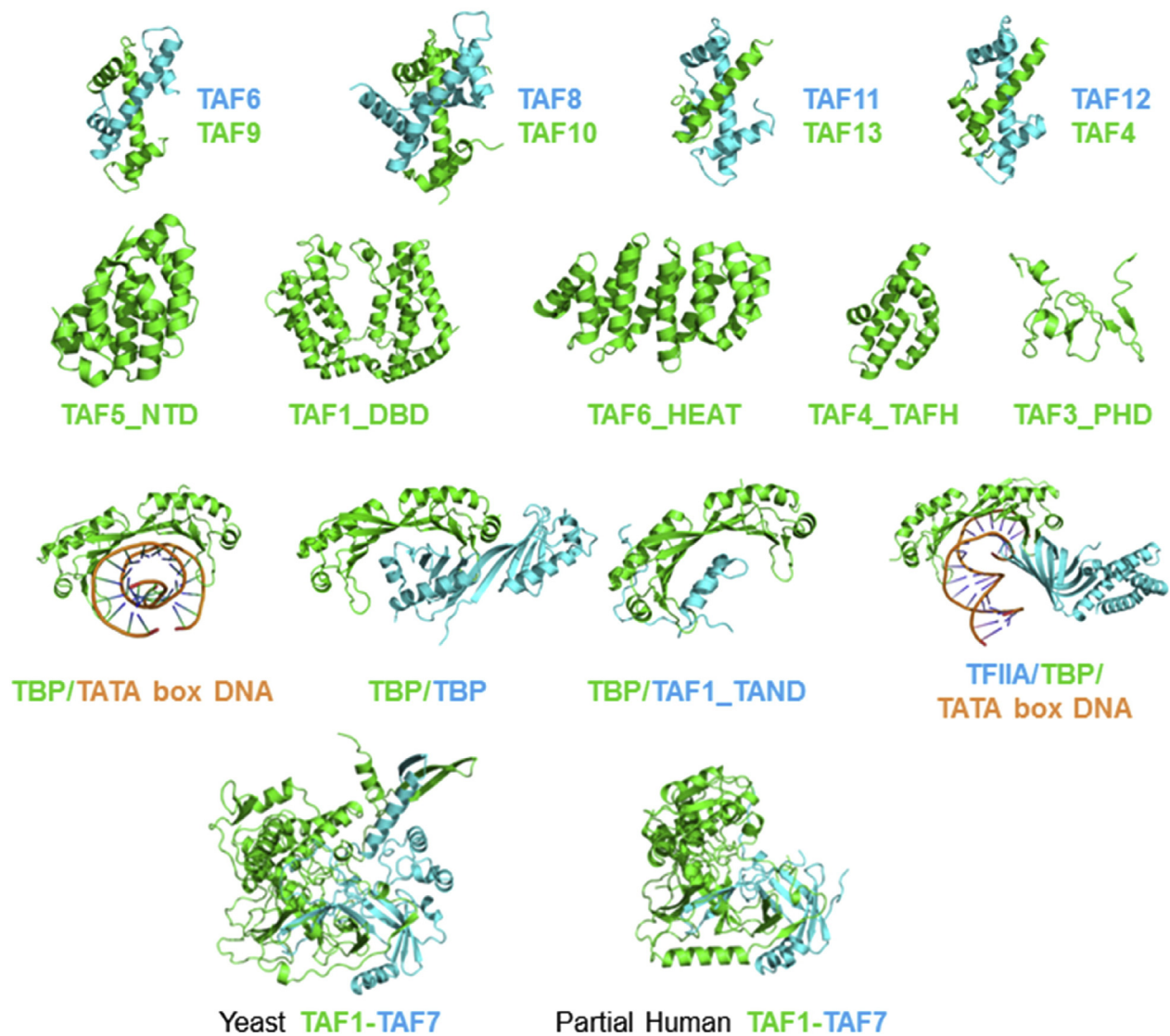


Fig. 5. TFIID Domain Structure Gallery. A range of structure of TFIID domains and domain components has been determined at near atomic resolution, mostly by X-ray crystallography. Structures shown here include structures of HFD pairs (top row): *Drosophila* TAF6–TAF9 (PDB ID: 1TAF) [54], human TAF8–TAF10 (PDB ID: 4WV4) [45], human TAF11–TAF13 (PDB ID: 1BH8) [52], and human TAF4–TAF12 (PDB ID: 1H3O) [53]. Crystal structures of isolated single domains of a variety of TAFs have been solved: human TAF5_NTD (PDB ID: 2NXP) [72], human TAF1_DBD (double bromodomain) (PDB ID: 1EQF) [73], the *A. locustae* TAF6_HEAT repeat (PDB ID: 4ATG) [74], human TAF4_TAFH (PDB ID: 2P6V) [75], and an NMR-based structure of mouse TAF3_PHD (PDB ID: 2K16) [76]. The conserved core of TBP has been studied intensively, and a selection of TBP-containing structures is shown: human TBP in complex with TATA-box DNA (PDB ID: 1CDW) [50], a yeast TBP dimer (PDB ID: 1TBP) [48], a yeast TBP–TAF1_TAND complex (PDB ID: 4B0A) [55], human TBP/TFIIA/DNA complex (PDB ID: 1NVP) [47]. More recently, crystal structures of pairwise interactions within TFIID other than HFDs have been obtained: a yeast TAF1–TAF7 complex (left) (PDB ID: 4OY2) and a partial human TAF1–TAF7 counterpart (right) (PDB ID: 4RGW) were crystallized and their structure determined [51].

A recent study provided a further piece to this puzzle by reporting the discovery of a novel TFIID submodule formed by TAF2, TAF8, and TAF10 in the cytoplasm of human cells (Fig. 4) [45]. This heterotrimeric complex was dissected by a combination of structural and biochemical methods, *in vitro* and *in vivo* [45]. TAF8 was found to form the backbone of this complex, mediating the interactions

with TAF10 by its HFD and with TAF2 by a previously uncharacterized domain in the C-terminal presumably unstructured region of TAF8 (Fig. 4). This TAF2-interaction domain did not only stabilize the TAF2–8–10 complex in the cytosol, but was also shown to be critical for TAF2 accretion into nuclear holo-TFIID, providing new insight into TFIID holo-complex formation. Furthermore, the TAF2–8–10

complex was shown to interact with importin- α via a nuclear localization signal present on TAF8, forming a putative nuclear import particle (NIP) (Fig. 4) [45]. Previous experiments had shown that the import of both TAF8 and TAF10 into the cell nucleus depends on the interaction of TAF8 with importin [46]. Now, it appears that TAF8 not only mediates co-import of TAF10, but also TAF2, by means of the importin α/β pathway and the formation of a TAF2–8–10–Importin NIP.

The existence of stable TFIID subcomplexes in the nucleus (e.g., core-TFIID) and in the cytosol (e.g. TAF2–8–10) supports the concept that preformed TFIID submodules exist, which then combine into the holo-complex, regulated by cellular processes such as nuclear import. It is tempting to speculate

that these subassemblies may have holo-TFIID-independent physiological functions, in transcription regulation but possibly also in other unrelated cellular processes. Further study is required to scrutinize and elucidate these exciting possibilities.

TFIID Domain Structure Gallery

Substantial research effort has been dedicated to uncover the molecular structure of TFIID, and much has been learned from the outcomes to date. X-ray crystallography has provided atomic structures of TFIID components, mostly conserved domain(s) in isolation or small complexes with other TAF domains, notably HFDs (Fig. 5). A considerable number of

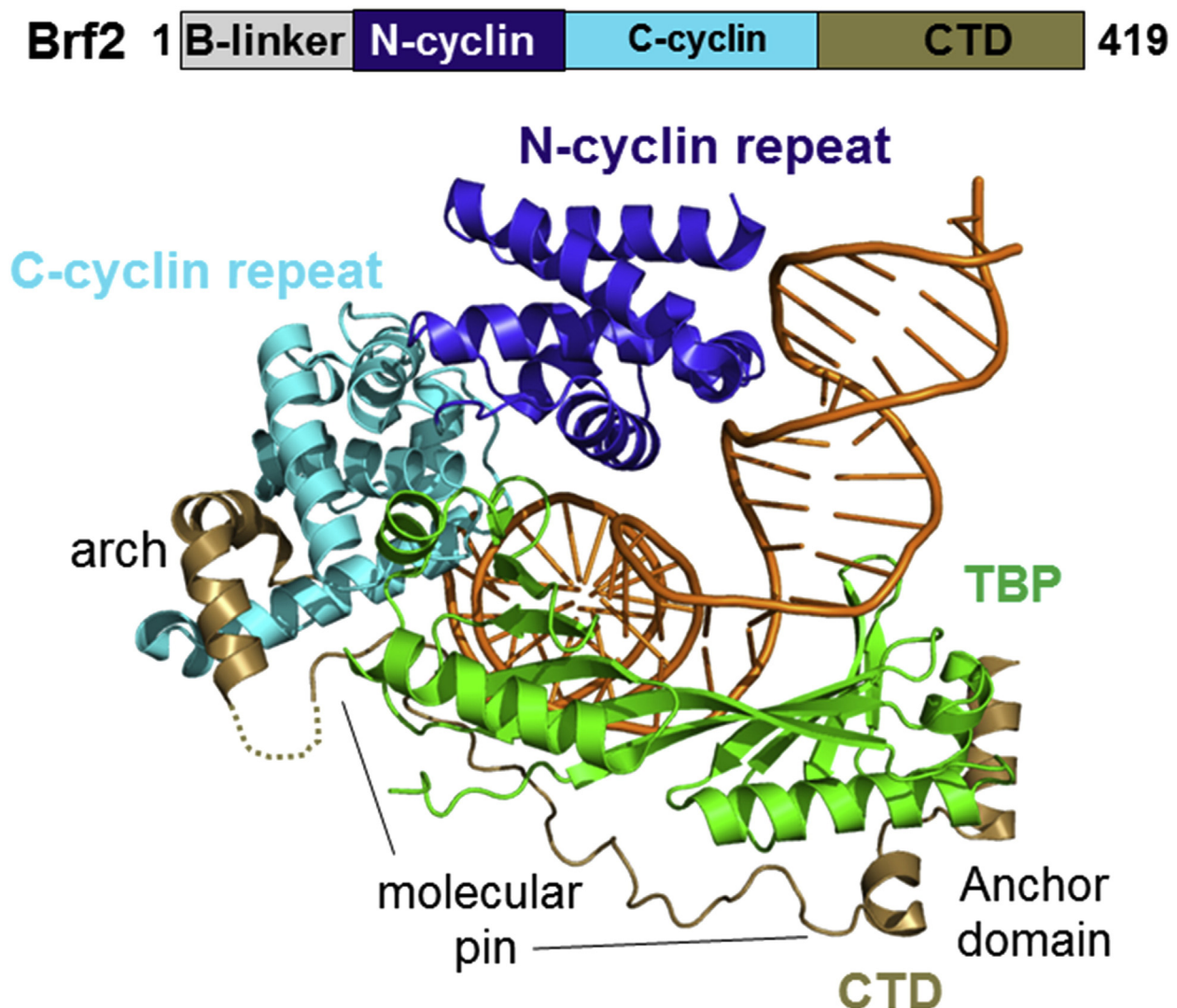


Fig. 6. Brf2–TBP–DNA Complex. The crystal structure of Brf2 bound to TBP (green) and promoter DNA (orange) is shown (PDB ID: 4ROC). Brf2 is shown in a schematic representation, depicted as a bar (top). Linker, cyclin domains, and CTD are indicated. The Brf2 CTD reaches over to interact with the convex surface of TBP, mediated by a molecular pin and an anchor domain (adapted from Ref. [67]).

structures of TBP exist, also bound to TFIIA, TFIIB, and TATA-box DNA [47–51]. Crystal structures of several HFD pairs in TFIID were solved [45,52–54]. NMR revealed an interesting case of TATA-box mimicry, with a low-complexity N-terminal domain of TAF1 occupying the DNA binding surface of TBP. This mimicry was later confirmed and expanded by X-ray crystallography [55–57]. Recently, a crystal structure of a TAF1–TAF7 complex from yeast [58] was determined, followed shortly by a partial structure of human TAF1–TAF7 complex [59]. These structures show that the TAF1–TAF7 interaction is intricate, involving a shared β -barrel motif between TAF1 and TAF7, which both donate β -strands that are intertwined. The β -barrel structure, conserved in yeast and human, is closely reminiscent of the β -barrel found in human TFIIF [60]. Interestingly, a recent study showed that a transcription factor, TFIIC, which is involved in regulating transcription by RNA pol III, also contains a similar β -barrel fold [61]. These novel findings indicate that this motif may be prevalent in transcription factor complexes and suggest an evolutionary conservation between TFIIF, TFIIC, and TFIID.

It is noteworthy that the TAF1 fragment used in the TAF1–TAF7 studies encompasses a region that was previously proposed to confer histone acetyltransferase (HAT) activity [62]. However, the corresponding fold in the TAF1–TAF7 crystal structure does not resemble a known HAT domain. It was suggested previously that TAF7 may inhibit the acetyltransferase activity upon binding to TAF1 [56]. Thus, it cannot be ruled out entirely that the binding of TAF7 to TAF1 reconfigures a functional HAT domain into the inactive fold observed in the TAF1–TAF7 model. Evidence suggests that TAF7 may be released from PIC after transcription initiation, when RNA pol II transits to transcript elongation [63,64]. Phosphorylation of TAF7 by a putative kinase activity contained within TAF1 was proposed to disrupt the TAF1–TAF7 interaction that is thought to result in the TAF7 release from PIC. Specific phosphorylation site(s) in TAF7 were found in its C-terminal part [64–66]. Interestingly, these are not present in the proposed interaction region between TAF1 and TAF7. Based on the crystal structure, it is unclear how the tight interactions in the TAF1–TAF7 complex would be disrupted by phosphorylation of TAF7, given that the proposed phosphorylation of TAF7 is not within the β -barrel region which would need to be unfolded.

New Insights from the RNA Pol III System

Recently, the structure of TFIIB-related factor 2 (Brf2) in complex with TBP and different natural promoters was determined, providing a detailed view of key interactions in RNA pol III transcription initiation (Fig. 6) [67]. This study validated the proposed structural and functional conservation

between TFIIB and TFIIB-related factors [68]. The overall architecture of Brf2–TBP–DNA is similar to the previously determined structures of TFIIB–TBP–DNA [69,70]. Brf2 is composed of N- and C-terminal cyclin repeats and a C-terminal domain (CTD). The cyclin repeats bind to the minor and major grooves of DNA, respectively, similar to the TFIIB/TBP/DNA ternary structure. The Brf2-specific CTD adopts three distinct structural elements: an “arch”, a “molecular pin” at the Brf2–TBP interface, and a “TBP anchor domain”, which reaches to the convex surface of TBP, reminiscent of TAF1–TBP and Brf1–TBP interactions [55,71]. Interestingly, in this study, a novel redox-sensing regulatory module was identified in Brf2, involved in redox-dependent control of RNA pol III transcription *in vivo*.

Conclusions

Unraveling the molecular mechanisms of transcription regulation has fascinated generations of researchers and this trend is unbroken. Advances notably in single particle cryo-EM technology are unlocking molecular structures of multiprotein transcription factors that are forming ever larger and more complex architectures. Crucial new insight about the supramolecular organization of key players, their structural plasticity and unexpected and pronounced dynamic rearrangements are emerging. Classical views of complexes existing as clearly defined, discrete, and uniform entities are challenged by the discovery of multiple isoforms and partial but physiological complexes. Complex assembly from preformed submodules is emerging as a key concept in transcription regulation. We have discussed just a few recent highlights in the present contribution. We anticipate many more thereof in the near future.

Acknowledgment

We thank all members of the Berger laboratory for helpful discussions. I.B. acknowledges support from the EMBL, the European Commission Framework Programme (FP) 7 ComplexINC project (contract no. 279039), the Agence National de Recherche (ANR) (contract number ANR-13-BSV8-0021) Projet Blanc DiscoverIID, and from the Wellcome Trust (WT106115AIA) through a Senior Investigator Award. M.H. was a Kekulé fellow of the Fonds der Chemischen Industrie (FCI, Germany). S.T. was supported by the European Commission through the Marie-Curie post-doctoral fellowship programme.

Received 4 February 2016;

Received in revised form 1 April 2016;

Accepted 1 April 2016

Available online xxxx

Keywords:

preinitiation complex;
general transcription factor TFIID;
transcription initiation;
RNA polymerase II;
multiprotein complex

Abbreviations used:

GTF, general transcription factors; RNA pol II, RNA polymerase II; PIC, preinitiation complex; TBP, TATA-box binding protein; ITC, initially transcribing complex; cryo-EM, cryo-electron microscopy; CLMS, crosslinking mass spectrometry; HFD, histone fold domain; SAGA, Spt-Ada-Gcn5-Acetyl transferase; NIP, nuclear import particle; HAT, histone acetyltransferase; Brf2, TFIIB-related factor 2; CTD, C-terminal domain; TAFs, TBP-associated factors.

References

- [1] B.J. Venters, B.F. Pugh, How eukaryotic genes are transcribed, *Crit. Rev. Biochem. Mol. Biol.* 44 (2009) 117–141.
- [2] T. Imasaki, G. Calero, G. Cai, K.L. Tsai, K. Yamada, F. Cardelli, et al., Architecture of the Mediator head module, *Nature*. 475 (2011) 240–243.
- [3] Y.J. Kim, S. Bjorklund, Y. Li, M.H. Sayre, R.D. Kornberg, A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II, *Cell*. 77 (1994) 599–608.
- [4] Z.C. Poss, C.C. Ebmeier, D.J. Taatjes, The Mediator complex and transcription regulation, *Crit. Rev. Biochem. Mol. Biol.* 48 (2013) 575–608.
- [5] M.C. Thomas, C.M. Chiang, The general transcription machinery and general cofactors, *Crit. Rev. Biochem. Mol. Biol.* 41 (2006) 105–178.
- [6] W. Wang, M. Carey, J.D. Gralla, Polymerase II promoter activation: closed complex formation and ATP-driven start site opening, *Science*. 255 (1992) 450–453.
- [7] T.K. Kim, R.H. Ebright, D. Reinberg, Mechanism of ATP-dependent promoter melting by transcription factor IIH, *Science*. 288 (2000) 1418–1422.
- [8] J. Fishburn, E. Tomko, E. Galbur, S. Hahn, Double-stranded DNA translocase activity of transcription factor TFIID and the mechanism of RNA polymerase II open complex formation, *Proc. Natl. Acad. Sci. U. S. A.* 112 (2015) 3961–3966.
- [9] F. Tirode, D. Busso, F. Coin, J.M. Egly, Reconstitution of the transcription factor TFIID: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7, *Mol. Cell*. 3 (1999) 87–95.
- [10] A. Dvir, S. Tan, J.W. Conaway, R.C. Conaway, Promoter escape by RNA polymerase II. formation of an escape-competent transcriptional intermediate is a prerequisite for exit of polymerase from the promoter, *J. Biol. Chem.* 272 (1997) 28,175–28,178.
- [11] A. Saunders, L.J. Core, J.T. Lis, Breaking barriers to transcription elongation, *Nat. Rev. Mol. Cell Biol.* 7 (2006) 557–567.
- [12] S. Hahn, Structure and mechanism of the RNA polymerase II transcription machinery, *Nat. Struct. Mol. Biol.* 11 (2004) 394–403.
- [13] K. Adelman, J.T. Lis, Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans, *Nat. Rev. Genet.* 13 (2012) 720–731.
- [14] H. Kwak, J.T. Lis, Control of transcriptional elongation, *Annu. Rev. Genet.* 47 (2013) 483–508.
- [15] N. Yudkovsky, J.A. Ranish, S. Hahn, A transcription reinitiation intermediate that is stabilized by activator, *Nature*. 408 (2000) 225–229.
- [16] D.S. Luse, G.A. Jacob, Abortive initiation by RNA polymerase II *in vitro* at the adenovirus 2 major late promoter, *J. Biol. Chem.* 262 (1987) 14,990–14,997.
- [17] F.C. Holstege, U. Fiedler, H.T. Timmers, Three transitions in the RNA polymerase II transcription complex during initiation, *EMBO J.* 16 (1997) 7468–7480.
- [18] Y. He, J. Fang, D.J. Taatjes, E. Nogales, Structural visualization of key steps in human transcription initiation, *Nature*. 495 (2013) 481–486.
- [19] K. Murakami, H. Elmlund, N. Kalisman, D.A. Bushnell, C.M. Adams, M. Azubel, et al., Architecture of an RNA polymerase II transcription pre-initiation complex, *Science*. 342 (2013) 1,238,724.
- [20] E. Kandiah, S. Trowitzsch, K. Gupta, M. Haffke, I. Berger, More pieces to the puzzle: recent structural insights into class II transcription initiation, *Curr. Opin. Struct. Biol.* 24 (2014) 91–97.
- [21] S. Sainsbury, J. Niesser, P. Cramer, Structure and function of the initially transcribing RNA polymerase II–TFIIB complex, *Nature*. 493 (2013) 437–440.
- [22] R. Kambadur, V. Culotta, D. Hamer, Cloned yeast and mammalian transcription factor TFIID gene products support basal but not activated metallothionein gene transcription, *Proc. Natl. Acad. Sci. U. S. A.* 87 (1990) 9168–9172.
- [23] B.D. Dynlacht, T. Hoey, R. Tjian, Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation, *Cell*. 66 (1991) 563–576.
- [24] C. Plaschka, L. Lariviere, L. Wenzek, M. Seizl, M. Hemann, D. Tegunov, et al., Architecture of the RNA polymerase II–Mediator core initiation complex, *Nature*. 518 (2015) 376–380.
- [25] L. Tora, A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription, *Genes Dev.* 16 (2002) 673–675.
- [26] S.L. Sanders, K.A. Garbett, P.A. Weil, Molecular characterization of *Saccharomyces cerevisiae* TFIID, *Mol. Cell. Biol.* 22 (2002) 6000–6013.
- [27] A. Hoffmann, C.M. Chiang, T. Oelgeschlager, X. Xie, S.K. Burley, Y. Nakatani, R.G. Roeder, A histone octamer-like structure within TFIID, *Nature*. 380 (1996) 356–359.
- [28] C. Leurent, S. Sanders, C. Ruhlmann, V. Mallouh, P.A. Weil, D.B. Kirschner, et al., Mapping histone fold TAFs within yeast TFIID, *EMBO J.* 21 (2002) 3424–3433.
- [29] F. Andel III, A.G. Ladurner, C. Inouye, R. Tjian, E. Nogales, Three-dimensional structure of the human TFIID-IIA-IIB complex, *Science*. 286 (1999) 2153–2156.
- [30] M. Brand, C. Leurent, V. Mallouh, L. Tora, P. Schultz, Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC, *Science*. 286 (1999) 2151–2153.
- [31] P. Grob, M.J. Cruse, C. Inouye, M. Peris, P.A. Penczek, R. Tjian, E. Nogales, Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues, *Structure*. 14 (2006) 511–520.
- [32] H. Elmlund, V. Baraznenok, T. Linder, Z. Szilagy, R. Rofougaran, A. Hofer, et al., Cryo-EM reveals promoter DNA binding and conformational flexibility of the general transcription factor TFIID, *Structure*. 17 (2009) 1442–1452.

- [33] M.A. Cianfrocco, G.A. Kassavetis, P. Grob, J. Fang, T. Juven-Gershon, J.T. Kadonaga, E. Nogales, Human TFIID binds to core promoter DNA in a reorganized structural state, *Cell*. 152 (2013) 120–131.
- [34] K.J. Wright, M.T. Marr II, R. Tjian, TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 12,347–12,352.
- [35] C. Bieniossek, G. Papai, C. Schaffitzel, F. Garzoni, M. Chaillet, E. Scheer, et al., The architecture of human general transcription factor TFIID core complex, *Nature*. 493 (2013) 699–702.
- [36] M.A. Demeny, E. Soutoglou, Z. Nagy, E. Scheer, A. Janoshazi, M. Richardot, et al., Identification of a small TAF complex and its role in the assembly of TAF-containing complexes, *PLoS One*. 2 (2007), e316.
- [37] F. Muller, L. Tora, The multicoloured world of promoter recognition complexes, *EMBO J.* 23 (2004) 2–8.
- [38] G.A. Maston, L.J. Zhu, L. Chamberlain, L. Lin, M. Fang, M.R. Green, Non-canonical TAF complexes regulate active promoters in human embryonic stem cells, *Elife*. 1 (2012), e00068.
- [39] R. Dikstein, S. Zhou, R. Tjian, Human TAFII 105 is a cell type-specific TFIID subunit related to hTAFII130, *Cell*. 87 (1996) 137–146.
- [40] J.C. Pointud, G. Mengus, S. Brancorsini, L. Monaco, M. Parvinen, P. Sassone-Corsi, I. Davidson, The intracellular localisation of TAF7L, a paralogue of transcription factor TFIID subunit TAF7, is developmentally regulated during male germ-cell differentiation, *J. Cell Sci.* 116 (2003) 1847–1858.
- [41] B. Bell, E. Scheer, L. Tora, Identification of hTAF(II)80 delta links apoptotic signaling pathways to transcription factor TFIID function, *Mol. Cell*. 8 (2001) 591–600.
- [42] X. Jacq, C. Brou, Y. Lutz, I. Davidson, P. Chambon, L. Tora, Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor, *Cell*. 79 (1994) 107–117.
- [43] R. van Nuland, A.W. Schram, F.M. van Schaik, P.W. Jansen, M. Vermeulen, H.T. Marc Timmers, Multivalent engagement of TFIID to nucleosomes, *PLoS One*. 8 (2013), e73495.
- [44] H.T. Timmers, L. Tora, SAGA unveiled, *Trends Biochem. Sci.* 30 (2005) 7–10.
- [45] S. Trowitzsch, C. Viola, E. Scheer, S. Conic, V. Chavant, M. Fournier, et al., Cytoplasmic TAF2–TAF8–TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules, *Nat. Commun.* 6 (2015) 6011.
- [46] E. Soutoglou, M.A. Demeny, E. Scheer, G. Fienga, P. Sassone-Corsi, L. Tora, The nuclear import of TAF10 is regulated by one of its three histone fold domain-containing interaction partners, *Mol. Cell Biol.* 25 (2005) 4092–4104.
- [47] M. Bleichenbacher, S. Tan, T.J. Richmond, Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes, *J. Mol. Biol.* 332 (2003) 783–793.
- [48] D.I. Chasman, K.M. Flaherty, P.A. Sharp, R.D. Kornberg, Crystal structure of yeast TATA-binding protein and model for interaction with DNA, *Proc. Natl. Acad. Sci. U. S. A.* 90 (1993) 8174–8178.
- [49] Y. Kim, J.H. Geiger, S. Hahn, P.B. Sigler, Crystal structure of a yeast TBP/TATA-box complex, *Nature*. 365 (1993) 512–520.
- [50] D.B. Nikolov, H. Chen, E.D. Halay, A. Hoffman, R.G. Roeder, S.K. Burley, Crystal structure of a human TATA box-binding protein/TATA element complex, *Proc. Natl. Acad. Sci. U. S. A.* 93 (1996) 4862–4867.
- [51] S. Tan, Y. Hunziker, D.F. Sargent, T.J. Richmond, Crystal structure of a yeast TFIIA/TBP/DNA complex, *Nature*. 381 (1996) 127–151.
- [52] C. Birck, O. Poch, C. Romier, M. Ruff, G. Mengus, A.C. Lavigne, et al., Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family, *Cell*. 94 (1998) 239–249.
- [53] S. Werten, A. Mitschler, C. Romier, Y.G. Gangloff, S. Thuault, I. Davidson, D. Moras, Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20), *J. Biol. Chem.* 277 (2002) 45,502–45,509.
- [54] Xie, X., Kokubo, T., Cohen, S. L., Mirza, U. A., Hoffmann, A., Chait, B. T., et al. Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature*. 380, 1996, 316–22.
- [55] M. Anandapadamanaban, C. Andresen, S. Helander, Y. Ohyama, M.I. Siponen, P. Lundstrom, et al., High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation, *Nat. Struct. Mol. Biol.* 20 (2013) 1008–1014.
- [56] D. Liu, R. Ishima, K.I. Tong, S. Bagby, T. Kokubo, D.R. Muhandiram, et al., Solution structure of a TBP–TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP, *Cell*. 94 (1998) 573–583.
- [57] T.K. Mal, Y. Masutomi, L. Zheng, Y. Nakata, H. Ohta, Y. Nakatani, et al., Structural and functional characterization on the interaction of yeast TFIID subunit TAF1 with TATA-binding protein, *J. Mol. Biol.* 339 (2004) 681–693.
- [58] S. Bhattacharya, X. Lou, P. Hwang, K.R. Rajashankar, X. Wang, J.A. Gustafsson, et al., Structural and functional insight into TAF1–TAF7, a subcomplex of transcription factor II D, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 9103–9108.
- [59] H. Wang, E.C. Curran, T.R. Hinds, E.H. Wang, N. Zheng, Crystal structure of a TAF1–TAF7 complex in human transcription factor IID reveals a promoter binding module, *Cell Res.* 24 (2014) 1433–1444.
- [60] F. Gaiser, S. Tan, T.J. Richmond, Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution, *J. Mol. Biol.* 302 (2000) 1119–1127.
- [61] N.M. Taylor, F. Baudin, G. von Scheven, C.W. Muller, RNA polymerase III-specific general transcription factor IIIC contains a heterodimer resembling TFIIF Rap30/Rap74, *Nucleic Acids Res.* 41 (2013) 9183–9196.
- [62] C.A. Mizzen, X.J. Yang, T. Kokubo, J.E. Brownell, A.J. Bannister, T. Owen-Hughes, et al., The TAF(II)250 subunit of TFIID has histone acetyltransferase activity, *Cell*. 87 (1996) 1261–1270.
- [63] A. Gegonne, B.N. Devaiah, D.S. Singer, TAF7: traffic controller in transcription initiation, *Transcription* 4 (2013) 29–33.
- [64] A. Gegonne, J.D. Weissman, M. Zhou, J.N. Brady, D.S. Singer, TAF7: a possible transcription initiation check-point regulator, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 602–607.
- [65] A. Gegonne, J.D. Weissman, D.S. Singer, TAFII55 binding to TAFII250 inhibits its acetyltransferase activity, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 12,432–12,437.
- [66] S.L. Kloet, J.L. Whiting, P. Gafken, J. Ranish, E.H. Wang, Phosphorylation-dependent regulation of cyclin D1 and cyclin A gene transcription by TFIID subunits TAF1 and TAF7, *Mol. Cell Biol.* 32 (2012) 3358–3369.
- [67] J. Gouge, K. Satia, N. Guthertz, M. Widya, A.J. Thompson, P. Cousin, et al., Redox signaling by the RNA polymerase III TFIIB-related factor Brf2, *Cell*. 163 (2015) 1375–1387.

- [68] A. Vannini, P. Cramer, Conservation between the RNA polymerase I, II, and III transcription initiation machineries, *Mol. Cell.* 45 (2012) 439–446.
- [69] D.B. Nikolov, H. Chen, E.D. Halay, A.A. Usheva, K. Hisatake, D.K. Lee, et al., Crystal structure of a TFIIB–TBP–TATA-element ternary complex, *Nature.* 377 (1995) 119–128.
- [70] F.T. Tsai, P.B. Sigler, Structural basis of preinitiation complex assembly on human pol II promoters, *EMBO J.* 19 (2000) 25–36.
- [71] Z.S. Juo, G.A. Kassavetis, J. Wang, E.P. Geiduschek, P.B. Sigler, Crystal structure of a transcription factor IIIB core interface ternary complex, *Nature.* 422 (2003) 534–539.
- [72] S. Bhattacharya, S. Takada, R.H. Jacobson, Structural analysis and dimerization potential of the human TAF5 subunit of TFIID, *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 1189–1194.
- [73] R.H. Jacobson, A.G. Ladurner, D.S. King, R. Tjian, Structure and function of a human TAFII250 double bromodomain module, *Science.* 288 (2000) 1422–1425.
- [74] E. Scheer, F. Delbac, L. Tora, D. Moras, C. Romier, TFIID TAF6–TAF9 complex formation involves the HEAT repeat-containing C-terminal domain of TAF6 and is modulated by TAF5 protein, *J. Biol. Chem.* 287 (2012) 27,580–27,592.
- [75] X. Wang, D.M. Truckses, S. Takada, T. Matsumura, N. Tanese, R.H. Jacobson, Conserved region I of human coactivator TAF4 binds to a short hydrophobic motif present in transcriptional regulators, *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 7839–7844.
- [76] H. van Ingen, F.M. van Schaik, H. Wienk, J. Ballering, H. Rehmann, A.C. Dechesne, et al., Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3, *Structure.* 16 (2008) 1245–1256.